

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開平10-240749

(43) 公開日 平成10年(1998)9月11日

(51) Int.Cl.^a

識別記号

F I
G O 6 F 15/403
15/40

340A
310F

審査請求 有 請求項の数 6 O.L. (全 11 頁)

(21) 出願番号 特願平9-38696

(22) 出願日 平成9年(1997)2月24日

(71)出願人 000004237
日本電気株式会社
東京都港区芝五丁目7番1号

(72)発明者 有吉 勇介
東京都港区芝五丁目7番1号 日本電気株
式会社内

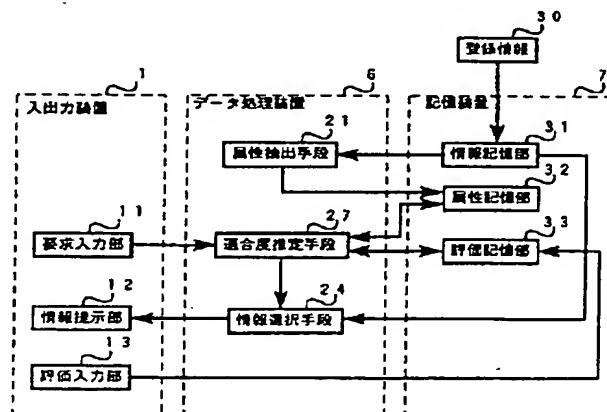
(74)代理人 弁理士 京本 直樹 (外2名)

(54) 【発明の名称】 情報フィルタリング方法及びその装置並びに情報フィルタリングプログラムを記録した記録媒体

(57) 【要約】 (修正有)

【課題】 従来のCBF (Constant Based Filtering) 方式では、要求者の評価のみから情報の属性と要求者の評価値の関係を学習し適合度を推定しており、従来のSIF (Social Information Filtering) 方式は、情報の属性を利用していないかったので、精度が低くなる欠点がある。

【解決手段】 要求者の評価と要求者以外の利用者の評価及び情報の属性を利用し要求者が評価していない情報に対する適合度を推定することにより、情報の推薦及びフィルタリングを行う。



1

【特許請求の範囲】

【請求項1】要求者に適合する情報の推薦やフィルタリングを行う情報フィルタリング方法であって、前記情報に含まれる属性を抽出して記憶するとともに、前記情報に対する利用者の評価を記憶しておき、前記要求者が評価した情報に対する評価と属性との関係と、前記要求者以外の利用者が評価した情報に対する評価と属性の関係を利用することによって、前記要求者が評価していない情報に対する前記要求者の適合度を推定し、

前記適合度を用いて前記要求者に対して適合する情報の推薦やフィルタリングを行うことを特徴とする情報フィルタリング方法。

【請求項2】要求者に適合する情報の推薦やフィルタリングを行う情報フィルタリング方法であって、前記情報に含まれる属性を抽出して属性記憶部に記憶するとともに、前記情報に対する利用者の評価値を受け付けて評価記憶部に記憶しておき、前記評価記憶部を参照して、前記要求者が評価した前記情報に対する評価値と、前記要求者以外の利用者が評価した前記情報に対する評価値から、前記要求者と前記要求者以外の利用者間の類似度を算出し、

算出された利用者間の類似度から、前記要求者以外の利用者が評価した情報のうち、前記要求者が評価していない情報について、前記要求者に対する第1の適合度を推定し、

前記属性記憶部と前記評価記憶部を参照して、前記要求者が評価を行った情報に含まれる属性と評価の関係を学習し、さらに、推定した前記第1の適合度から、前記要求者以外の利用者が評価した情報のうち、前記要求者が評価していない情報に含まれる属性と適合度の関係を学習し、

学習した前記属性と評価の関係及び前記属性と適合度の関係から、前記要求者と前記要求者以外の利用者が評価していない情報についての前記要求者の第2の適合度を推定し、

推定した前記第1の適合度と前記第2の適合度を統合することによって、前記要求者に対して適合する情報の推薦やフィルタリングを行うことを特徴とする情報フィルタリング方法。

【請求項3】要求者に適合する情報の推薦やフィルタリングを行う情報フィルタリング装置であって、前記情報に含まれる属性を抽出して記憶する属性記憶部と、前記情報に対する利用者の評価を記憶する評価記憶部と、前記要求者が評価した情報に対する評価と属性との関係と、前記要求者以外の利用者が評価した情報に対する評価と属性の関係を利用することによって、前記要求者が評価していない情報に対する前記要求者の適合度を推定し、

2

する適合度推定手段と、を備え、

推定された前記適合度を用いて前記要求者に対して適合する情報の推薦やフィルタリングを行うことを特徴とする情報フィルタリング装置。

【請求項4】要求者に適合する情報の推薦やフィルタリングを行う情報フィルタリング装置であって、前記情報に含まれる属性を抽出する属性抽出手段と、抽出された前記属性を記憶する属性記憶部と、複数の利用者から前記情報に対する評価値の入力を受け付ける評価入力部と、

前記評価入力部で受け付けた情報の評価値を記憶する評価記憶部と、

前記評価記憶部を参照して、前記要求者が前記情報に対して入力した評価値と、前記要求者以外の利用者が前記情報に対して入力した評価値から、前記要求者と前記要求者以外の利用者間の類似度を算出する利用者間類似度算出手段と、

前記利用者間類似度算出手段で算出された類似度から、前記要求者以外の利用者が評価した情報のうち、前記要求者が評価していない情報について、前記要求者に対する第1の適合度を推定する第1の適合度推定部と、

前記属性記憶部と前記評価記憶部を参照して、前記要求者が評価を行った情報に含まれる属性と評価の関係を学習するとともに、前記第1の適合度推定部で推定された第1の適合度を用いて、前記要求者以外の利用者が評価した情報のうち、前記要求者が評価していない情報に含まれる属性と適合度の関係を学習する利用者プロファイル学習手段と、

前記利用者プロファイル学習手段が学習した属性と評価の関係及び属性と適合度の関係から、前記要求者と前記要求者以外の利用者が評価していない情報について、前記要求者に対する第2の適合度を推定する第2の適合度推定部と、前記第1の適合度と前記第2の適合度を統合することによって、前記要求者に対して適合する情報の推薦やフィルタリングを行う適合度統合手段と、を備えることを特徴とする情報フィルタリング装置。

【請求項5】コンピュータによって要求者に適合する情報の推薦やフィルタリングを行う情報フィルタリングプログラムを記録した記録媒体において、

前記コンピュータの記憶装置に、前記情報に含まれる属性を抽出して記憶するとともに、前記情報に対する利用者の評価を記憶しておき、

前記要求者が評価した情報に対する評価と属性との関係と、前記要求者以外の利用者が評価した情報に対する評価と属性の関係を利用することによって、前記要求者が評価していない情報に対する前記要求者の適合度を推定し、

前記適合度を用いて前記要求者に対して適合する情報の推薦やフィルタリングを行う機能を有するプログラムを記録したことを特徴とする情報フィルタリングプログラ

50

ムを記録した記録媒体。

【請求項6】コンピュータによって要求者に適合する情報の推薦やフィルタリングを行う情報フィルタリングプログラムを記録した記録媒体において、コンピュータを、

前記情報に含まれる属性を抽出する属性抽出手段と、コンピュータの記憶手段に前記属性抽出手段によって抽出された前記属性を記憶させる属性記憶部と、コンピュータの記憶手段に、コンピュータの入力手段によって複数の利用者から前記情報に対する評価値の入力を受け付け、受け付けた情報の評価値を記憶させる評価記憶部と、

前記評価記憶部を参照して、前記要求者が前記情報に対して入力した評価値と、前記要求者以外の利用者が前記情報に対して入力した評価値から、前記要求者と前記要求者以外の利用者間の類似度を算出する利用者間類似度算出手段と、

前記利用者間類似度算出手段で算出された類似度から、前記要求者以外の利用者が評価した情報のうち、前記要求者が評価していない情報について、前記要求者に対する第1の適合度を推定する第1の適合度推定部と、前記属性記憶部と前記評価記憶部を参照して、前記要求者が評価を行った情報に含まれる属性と評価の関係を学習するとともに、前記第1の適合度推定部で推定された第1の適合度を用いて、前記要求者以外の利用者が評価した情報のうち、前記要求者が評価していない情報に含まれる属性と適合度の関係を学習する利用者プロファイル学習手段と、

前記利用者プロファイル学習手段が学習した属性と評価の関係及び属性と適合度の関係から、前記要求者と前記要求者以外の利用者が評価していない情報について、前記要求者に対する第2の適合度を推定する第2の適合度推定部と、

前記第1の適合度と前記第2の適合度を統合することによって、前記要求者に対して適合する情報の推薦やフィルタリングを行う適合度統合手段と、して機能させるためのプログラムを記録したことを特徴とする情報フィルタリングプログラムを記録した記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明はデータベースやWWW (World Wide Web)などの情報システムから、利用者の興味や嗜好に合った情報の推薦やフィルタリングを行う情報フィルタリング方法及び情報フィルタリング装置並びに情報フィルタリングプログラムを記録した記録媒体に関する。

【0002】

【従来の技術】従来、この種の情報フィルタリング方法や装置や情報フィルタリングプログラム（以下、まとめて情報フィルタリング方式と記載する。）はデータベー

スやパソコン通信やWWWなどの情報システムが提供する大量の情報や多数のサービス（以下、本明細書において、単に「情報」と記載した場合は、このことを指す。）の中から利用者の興味や嗜好に適合するものを推薦したりフィルタリングを行うために用いられている。

【0003】従来の情報フィルタリング方式には、CBF (Constant Based Filtering) 方式と、SIF (Social Information Filtering) 方式の2種類の方式が知られている。

10 【0004】このCBF方式は、情報の中に含まれているキーワードや単語頻度などデータ（本明細書では、このデータのことを属性と記載する。）から、各情報の内容の特徴を表現し、利用者の過去の興味などをキーワードの重みなどで表現したものとのマッチングに基づいて情報のフィルタリングを行う方式である。

【0005】また、SIF方式は、利用者が個々の情報に対して評価を行ったデータ（以下、本明細書では、評価と記載する。）を記憶しておき、自分（情報フィルタリングの要求者）が過去に行った評価と、要求者以外の利用者が過去に行った評価を比較して、要求者の嗜好に似た利用者を見つけだし、その嗜好の似た利用者が良い評価を受けた、要求者がまだ見ていない情報を選び出して要求者に情報のフィルタリングを行う方式である。

【0006】このCBF方式の情報フィルタリング方式の従来の技術の公知例としては、特開平3-94375号公報「文書検索装置」が知られており、一方、SIF方式の例としてはProceedings of the CSCW 1994 (proceedings of the cscw 1994) 1994年エイシーエムプレス (acm press) 175頁から186頁米国でGroupLensというシステムについて、Proceedings of the CHI 95 (proceedings of the chi 95) 1995年エイシーエムプレス (acm press) 210頁から217頁に記載の米国のRingoというシステムが知られている。

【0007】次に、CBF方式、SIF方式の従来の技術を図面を参照して説明する。

【0008】まず、CBF方式について説明する。CBF方式の従来技術を、図面の図5、図6を参照して説明する。図5を参照すると、CBF方式の従来の技術においては、入出力装置1とデータ処理装置2と、記憶装置3とを含んで構成されている。

【0009】入出力装置1は要求入力部11と、情報提示部12と、評価入力部13とを備える。

【0010】記憶装置3は情報記憶部31と、属性記憶部32と評価記憶部33を備えている。情報記憶部31は、登録情報30をあらかじめ記憶している。評価記憶部33には、情報に対する利用者の評価値があらかじめ記憶されている。

【0011】データ処理装置2は、属性抽出手段21と、利用者プロファイル学習手段22と、適合度推定手段23と情報選択手段24とを備える。

【0012】属性抽出手段21は情報記憶部31に記憶された情報から、特徴となる属性を抽出し属性記憶部32に格納する。また、ここでいう属性とは、先に説明した通り、情報の内容をキーワードや単語頻度などから特徴づけたものである。例えば、「ある情報の中には、

「野球」と「サッカー」いうキーワードが含まれる。」というもので、情報の特徴を抽出したものである。

【0013】利用者プロファイル学習手段22は、要求入力部11より要求者からの推薦要求を受けとると、評価記憶部33に記憶された、要求者が過去に行った情報に対する評価と、属性記憶部32に記憶された情報の属性のうち、要求者が評価済みの情報の属性から、要求者の評価と属性の間の関係を表すプロファイルを学習する。このプロファイルの学習は、例えば、要求者は、過去に「野球」というキーワードを含む情報に対して高い評価を行っていたが、逆に「サッカー」というキーワードを含む情報に対しては、低い評価を行っていたというように、要求者の評価と属性との間の関係が分かるものであり、言い換えれば、利用者興味を情報に含まれる単語（キーワード）の重みで表現したものである。

【0014】適合度推定手段23は利用者プロファイル学習手段22が学習した要求者のプロファイルと、属性記憶部32が記憶している情報の属性から、各情報の要求者に対する適合度を推定する。次に、情報選択手段24は、適合度推定手段23が推定した適合度を利用して、要求者に適合する情報を情報記憶部に記憶している情報の中から選択し情報提示部12に出力する。この例では、適合度推定手段23が、要求者が「野球」の属性を持った情報に対して興味があるというプロファイルを学習し、逆に「サッカー」の属性を持った情報に対しては興味がないというプロファイルを学習した結果を受け取り、要求者には、「野球」というキーワードを持った情報が適合すると推定する。情報選択手段24は、適合度推定手段の推定結果を受け取り、「野球」という属性を持った情報のマッチングを行い情報記憶部の中から過去に要求者が評価を行っていない情報を選択して情報提示部12に出力する。

【0015】要求者は情報提示部12で提示された情報が自分の要求に合うか評価する。評価入力部13から入*

$$\text{プロファイル} = \frac{\sum_{k=1}^n (\text{正規化された} k\text{番目の情報に対する評価値} \times k\text{番目の情報の特徴ベクトル})}{\sum_k (\text{正規化された} k\text{番目の情報に対する評価値})} \quad [\text{式1}]$$

【0022】また、情報の種類によっては特徴ベクトルの各要素値として情報の属性値そのままではなく、属性値の平均からの偏差を標準偏差で割った値を用いる方が精度が良い場合もある。

*力された評価は評価記憶部33に記憶される。よって、このことにより、適合フィードバックが行われ、利用者の興味にあった情報に含まれる単語（キーワード）の重要度を上げ、興味に合わなかった情報に含まれる単語の重要度は下げることができ、より利用者の興味に合った情報ファイルタリングを行うことが可能である。

【0016】次に、従来のCBF方式の動作について図6を参照して説明する。

【0017】要求入力部11から推薦要求が入力されると、利用者プロファイル学習手段22は、評価記憶部33に記憶された要求者の評価履歴と、属性記憶部31に記憶された情報の属性のうち、要求者が評価済みの情報の属性から、要求者の評価と属性の間の関係を表すプロファイルを学習する（C1）。続いて適合度推定手段23は利用者プロファイル学習手段22が学習した要求者のプロファイルと、属性記憶部32が記憶している情報の属性から、各情報の要求者に対する適合度を推定する（C2）。情報選択手段24は、適合度推定手段23が推定した適合度を利用して、要求者の要求に合う情報を選択し（C3）結果を情報提示部12に出力する（C4）。

【0018】情報提示部12に提示された情報に対して、要求者は自分の要求に合うか評価し、評価結果を評価入力部13から入力する。入力された評価情報は評価記憶部33に記憶される。

【0019】よって、上述したCBF方式の情報ファイルタリングでは、情報の中に含まれているキーワードや単語頻度などの属性から、各情報の内容の特徴を表現し、利用者の過去の興味をキーワードなどで表現したものとのマッチングに基づいて情報のファイルタリングを行うことが可能である。

【0020】また、この説明に使用した例は、簡単に説明するための例であり、実際のプロファイル学習の方法としては、例えば次のことが考えられる。情報の特徴ベクトルとして情報の属性の値を並べたものを考え、情報の評価値は1から-1の間に正規化してあるものとする。ただし正規化された評価値は1が利用者の要求に合っている、-1が要求に合っていないものとする。このときプロファイルは式1で計算することができる。

【0021】

【数1】

【0023】さらに、情報の特徴ベクトルをそれぞれのベクトルの絶対値で割ることによりベクトルの長さを1に正規化した方が精度が良い場合もある。

【0024】適合度の推定方法としては情報の特徴ベ

クトルとプロファイルとのピアソンの積率相関係数を使うことが考えられる。

【0025】次に、SIF方式について説明する。ここでは、SIF方式の従来の技術を、図面の図7、図8を参照して簡単に説明する。また、前述したCBF方式の従来の技術と同様な機能をもつ構成については、説明を省略する。

【0026】図7を参照すると、この従来のSIF方式は、出入力装置1とデータ処理装置4と、記憶装置5とを含む。出入力装置1は要求入力部11と、情報提示部12と、評価入力部13とを備える。

【0027】記憶装置5は情報記憶部31と、評価記憶部33とを備えている。情報記憶部31は、登録情報30をあらかじめ記憶している。評価記憶部33は、利用者による情報に対する評価値をあらかじめ記憶している。この評価値は、情報フィルタリングの要求者が過去に評価を行ったものと、要求者以外の利用者がある情報に対して評価を行ったものについて保持している。

【0028】データ処理装置4は、利用者間類似度算出手段25と、適合度推定手段26と、情報選択手段24とを備える。

【0029】利用者間類似度算出手段25は、要求入力部11で要求者からの推薦要求を受けとると、評価記憶部33に記憶された要求者の情報に対する評価履歴と要求者以外の利用者の評価履歴から利用者間の類似度を算出する。つまり、自分(要求者)と似た嗜好を持つ別の利用者を見つけだす。ただし、自分の嗜好に似た別の利用者は1人でなくても、複数でもよい。

【0030】適合度推定手段26は、評価記憶部33が記憶している評価と、利用者間類似度算出手段25が算出した利用者間類似度から、各情報の要求者に対する適合度を推定し、情報選択手段24は、適合度推定手段26が推定した適合度を利用して、要求者の要求に合う情報を選択し情報提示部12に出力する。つまり、利用者間類似度算出手段25が算出した利用者間類似度から、自分の嗜好に似た利用者が良い評価を受けた要求者がま*

$$\text{適合度} = \frac{\sum_k (\text{要求者と } k\text{ 番目の利用者の類似度} \times k\text{ 番目の利用者の評価値})}{\sum_k (\text{要求者と } k\text{ 番目の利用者の類似度})} \quad [\text{式2}]$$

【0037】情報提示部12に提示された情報に対して、要求者は自分の要求に合うか評価し、評価結果を評価入力部13から入力する。入力された評価情報は評価記憶部33に記憶される。

【0038】よって、SIF方式では、要求者が過去に行った評価と、要求者以外の利用者が過去に行った評価を比較して、要求者の嗜好に似た利用者を見つけだし、その嗜好の似た利用者が良い評価を受けた、要求者がまだ見ていらない情報を選び出して要求者に対して情報のフィルタリングを行うことができる。

*まだ見ていらない情報を適合度推定手段26で適合度を推定することによって、情報選択手段24を用いて選択して情報提示部12に提示する。

【0031】要求者は情報提示部12で提示された情報が自分の要求に合うか評価する。評価入力部13から入力された評価は評価記憶部33に記憶される。

【0032】このSIF方式を例を挙げて説明する。

「野球」について興味があるが、「サッカー」については、興味のない要求者に情報フィルタリングすることを前提に考える。このSIF方式では、要求者と同じように「野球」に興味があるが、「サッカー」に興味のない利用者を探しだし、その探し出された利用者が高い評価をした情報について要求者に提供する。よって、探し出された利用者が、「野球」の他に「釣り」に興味があれば、要求者に対して「釣り」の情報が提供されることになる。もちろん、提供される情報は、要求者がまだ見ていない情報であることが望ましい。

【0033】次に、図8を参照して、従来のSIF方式の動作について説明する。

【0034】要求入力部11から推薦要求があると、利用者間類似度算出手段25は評価記憶部33に記憶されている評価を利用して利用者間類似度を算出する(D1)。

【0035】例えば前述したGroupLensとRingoでは利用者間類似度としてピアソンの積率相関係数を使っている。続いて適合度推定手段26は、評価記憶部33が記憶している評価と、利用者間類似度算出手段25が算出した利用者間類似度から、各情報の要求者に対する適合度を推定する(D2)。例えば前述したGroupLensでは適合度を次にあげる式2により推定している。情報選択手段24は、適合度推定手段26が推定した適合度を利用して、要求者の要求に合う情報を選択し(D3)、結果を情報提示部12に出力する(D4)。

【0036】

【数2】

【0039】

【発明が解決しようとする課題】次に、この従来の技術の問題点を図面の図9を用いて説明する。

【0040】上述したCBF方式では、自分が評価した情報(A+C)に対して単語頻度等の情報の属性と、自分の評価との関係を学習し、自分が未評価の情報(B+D)をフィルタリングする。一方、SIF方式では、自分が評価した情報と、さらに他の利用者も評価した情報(C)について、自分と他の利用者の評価から自分と他の利用者の関係を学習し、自分は評価していないが、他

の利用者が評価した情報（B）をフィルタリングする。
【0041】つまり、従来のCBF方式の第1の問題点は、利用者が興味を持っている話題（つまり、キーワードなどの属性）についての情報を得ることが出来るが、情報に価値があるかは結局利用者が読んで判断しないといけないことである。その理由は、CBF方式はキーワードなどの属性や単語出現頻度に基づいた方式で、SIF方式のように人の評価に基づいていないからである。

【0042】従来のCBF方式の第2の問題点は、未知の単語や未知の属性値を持つ情報に対するフィルタリングの精度が低いことである。その理由はCBF方式は、要求者の評価履歴と要求者が評価済みの情報の属性から、要求者の評価と属性の間の関係をプロファイルとして学習し、そのプロファイルに基づいて適合度を推定するからである。そのため、未知の単語、属性値を持っていました場合、それを適合度にどのように反映すれば良いかCBF方式では分からぬという問題がある。

【0043】次に従来のSIF方式の第1の問題点は、利用者のうちの誰かが推薦・評価した情報しか得られないことである。その理由は、SIF方式が利用者の推薦と評価に基づいた方法で評価・推薦された情報しかフィルタリングの対象にならないためである。

【0044】従来のSIF方式の第2の問題点は、ある程度の量の評価が集まらなければフィルタリングの精度が低いことである。この理由はSIF方式が利用者の推薦と評価に基づいた方式だからである。

【0045】よって、本発明の目的は、これらの従来のCBF方式、SIF方式の問題を軽減または解消し、これらの従来の技術よりフィルタリング精度が高く、より利用者の興味や嗜好に合った情報の推薦やフィルタリングを行う情報フィルタリング方式を提供することにある。

【0046】

【課題を解決するための手段】従来のCBF方式では要求者の過去の情報に対する評価から、情報の属性と要求者の評価値の関係を学習し適合度を推定しており、要求者以外の評価を利用ていなかった。また、従来のSIF方式は要求者と要求者以外の利用者の評価を利用して適合度を推定しており、情報の属性を利用ていなかった。

【0047】それのことに対し、本発明の請求項1から6に記載された第1の発明は、要求者の評価と要求者以外の利用者の評価と情報の属性を利用し、適合度を推定することを特徴としている。よって、第1の発明は、利用者の評価を利用してることにより、前述したCBF方式の第1の問題点を軽減する。また、第1の発明は要求者の評価した情報に無い属性値や未知語に対しても、他の利用者が評価した情報に含まれていれば、適合度の推定に利用できるので、CBF方式の第2の問題点が軽減する。

【0048】また、第1の発明は、情報の属性も利用して適合度を推定しているので、誰も推薦・評価していない情報であっても従来のCBF方式の精度で適合度を推定できる。そのため前述したSIF方式の第1の問題と第2の問題を解消している。

【0049】つまり、本発明の第1の発明の特徴を図9を利用して説明すると、図9の（B）で示される領域では、SIF方式の方が上述した特徴をもつため、SIF方式を利用する。また、領域（D）については、情報フィルタリングの対象としているのは、CBF方式だけであるので、（D）では、CBF方式を用いる。ただし、CBF方式は前述したように（D）においても、未知語に弱いという問題点があるので、これをSIF方式のフィルタリング結果をCBF方式のプロファイルにフィードバックすることで軽減している。よって、本発明では、誰も評価していない情報（D）についても、情報フィルタリングを行うことが可能であり、さらに自分以外の利用者の評価を利用することにより従来のCBF方式より未知語に強いことが特徴である。

【0050】

【発明の実施の形態】次に、本発明の実施の形態について図面を参照して詳しく説明する。また、この発明の実施の形態においては、上述した従来の技術と同様な構成については、図面に同一番号を付けている。特に説明が無い限り、従来の技術と同様な構成であるため説明を省略する。

【0051】まず、本発明の第1の発明の実施の形態について図面を参照して説明する。

【0052】図1を参照すると、第1の発明の実施の形態は入出力装置1とデータ処理装置6と、記憶装置7とを含む。

【0053】入出力装置1は要求入力部11と、情報提示部12と、評価入力部13とを備えている。

【0054】記憶装置7は情報記憶部31と、属性記憶部32と評価記憶部33を備えている。情報記憶部31は、登録情報30をあらかじめ記憶している。評価記憶部33は、情報に対する利用者の評価値を、要求者および要求者以外の利用者による評価値も含めてあらかじめ記憶している。

【0055】データ処理装置6は、属性抽出手段21と、適合度推定手段27と、情報選択手段24を備えている。

【0056】属性抽出手段21は情報記憶部31に記憶された情報から特徴となる属性（例えば、キーワード）を抽出し属性記憶部32に格納する。

【0057】適合度推定手段27は、要求入力部11で要求者からの推薦要求を受けとると、評価記憶部33に記憶された要求者の評価と要求者以外の利用者の評価（つまり図9の（C））を利用して利用者間の類似度を算出して、まず図9の（B）における情報の適合度を算

11

出して情報フィルタリングを行う。

【0058】次に、図9の(D)に対する情報フィルタリングを行うが、従来のCBF方式においては、誰も評価をしていない情報である(D)の領域に対して、自分が過去に評価した情報(つまり図9の(A+C))の属性と評価の関係を図5の利用者プロファイル学習手段22で学習し、情報フィルタリングを行っていた。しかし、本発明においては、先に行われた(B)に対する情報フィルタリングの結果も利用し、フィルタリングされた(B)の情報の属性と、要求者に対する適合度(本明細書では、適合度と評価は同じ意味を持つ。ただし、利用者が入力したものと評価と記載し、計算により要求者の適合の度合いを推定したものと適合度と、場合分けして記載する。)の関係も学習する。この学習結果により、誰も評価していない情報(D)については、(A+B+C)に含まれる情報の属性と評価(もしくは適合度)の関係を学習して、情報フィルタリングを行う。

【0059】情報選択手段24は、適合度推定手段27の情報フィルタリング結果を情報提示部12に出力する。要求者は情報提示部12で提示された情報が自分の要求に合うか評価する。

【0060】評価入力部13から入力された情報に対する評価は評価記憶手段33に格納される。

【0061】次に、図2を参照して、第1の発明の動作の実施の形態について説明する。

【0062】要求入力部11から要求者が推薦要求を入力すると、適合度推定手段27は、評価記憶部33に記憶された要求者の評価と要求者以外の利用者の評価と、属性記憶部32に記憶された情報の属性を利用して、利用者の評価値と情報の属性の間の関係を分析し、情報の要求者に対する適合度を推定する(A1)。情報選択手段24は、適合度推定手段27が推定した適合度を利用して、要求者の要求に合う情報を選択し(A2)、結果を情報提示部12に出力する(A3)。

【0063】次に、第1の発明の適合度推定手段27のより具体的な実施の形態について、図面を参照して詳細に説明する。

【0064】適合度推定手段27は、利用者間類似度算出手段271と第1の適合度推定部272と利用者プロファイル学習手段273と第2の適合度推定部274と適合度統合手段275を備える。

【0065】利用者間類似度算出手段271は、要求入力部11で要求者からの推薦要求を受けとると、評価記憶部33に記憶された情報に対する要求者の評価と要求者以外の利用者の評価を利用して利用者間の類似度を算出する。

【0066】第1の適合度推定部272は、評価記憶部33が記憶している要求者以外の利用者の評価と、利用者間類似度算出手段271が算出した利用者間類似度から、図9の(B)で示される要求者以外の利用者が評価

12

した情報の要求者に対する適合度を推定する。

【0067】利用者プロファイル学習手段273は、属性記憶部32に記憶された情報の属性と評価記憶部33に記憶された要求者の評価と第1の適合度推定部272で推定された適合度から、情報の属性と要求者の評価との関係をあらわすプロファイルを学習するとともに、情報の属性と要求者に対する適合度との関係をあらわすプロファイルについても学習する。また、ここで言うプロファイルとは、利用者興味をキーワードなどの重みなどで表現したものである。

【0068】また、情報の属性と要求者の評価との関係をあらわすプロファイルとは、ある属性が付いた情報に対して、過去に要求者(利用者)がどのような評価を付けたかという関係を表現したものである。例えば、「要求者は、過去に「野球」というキーワードが含む情報について高い評価をしている。」という属性と要求者との関係を表現したものである。

【0069】次に、情報の属性と要求者に対する適合度との関係をあらわすプロファイルとは、第1の適合度推定部272で推定された各情報に対する要求者の適合度から、属性記憶部32を参照することによって属性と要求者に対する適合度との関係を表現したものである。

【0070】例えば、要求者は「野球」というキーワードが含まれる情報に対して過去に高い評価をしていたことを考える。この場合、要求者と同様に「野球」というキーワードを含む情報について高い評価をしている要求者以外の利用者が、要求者以外の利用者は評価しているが、要求者がまだ評価していない情報(つまり、図9の(B)で示される情報)について評価した情報を取り出し、この情報に含まれる属性を抽出する。例えば、要求者と似た嗜好を持つ別の利用者が、「野球」の他に「釣り」というキーワードを持つ情報に対して高い評価をしている場合は、「要求者には、「野球」の他に「釣り」というキーワードを持つ情報について適合度が高そうだ。」という、属性と要求者に対する適合度の関係をあらわすプロファイルを学習する。

【0071】次に、第2の適合度推定部274は属性記憶部32が記憶している情報の属性と、利用者プロファイル学習手段273が学習した要求者のプロファイルから、利用者の誰も評価していない情報(つまり、図9の(D)で示される情報)の要求者に対する適合度を算出し、適合度統合手段275に伝える。つまり、この例の場合、図9の(D)で示される情報中に、「野球」というキーワードを含む情報については算出される適合度は高くなるが、同様に「釣り」というキーワードを持つ情報について算出される適合度は高くなる。

【0072】また、適合度推定手段274は、誰も評価していない情報に加えて評価した利用者数が少ない情報の適合度も推定することも考えられる。

【0073】適合度統合手段275は第1の適合度推定

50

13

部272と第2の適合度推定部274から算出される適合度を統合して、情報選択手段24に伝える。

【0074】次に図3の実施の形態の動作を図4を参照して説明する。

【0075】要求入力部11から要求者からの推薦要求を受けとると、利用者間類似度算出手段271は、評価記憶部33に記憶された要求者と要求者以外の利用者の評価を利用して利用者間の類似度を算出する(A11)。第1の適合度推定部272は、評価記憶部33が記憶している要求者以外の利用者の評価と、利用者間類似度算出手段271が算出した利用者間類似度から、各情報の要求者に対する適合度を推定する(A12)。

【0076】利用者プロファイル学習手段273は、属性記憶部32に記憶された情報の属性と評価記憶部33に記憶された要求者の評価と第1の適合度推定部272で推定された適合度から、情報の属性と要求者の評価の関係をあらわすプロファイルと、情報の属性と要求者に対する適合度との関係をあらわすプロファイルを学習する(A13)。

【0077】第2の適合度推定部274は属性記憶部32が記憶している情報の属性と、利用者プロファイル学習手段273が学習した要求者のプロファイルから、誰も評価していない各情報の要求者に対する適合度を推定し、適合度統合手段275に伝える(A14)。このとき、誰も評価していない情報に加えて評価した利用者数が少ない情報の要求者に対する適合度を推定しても良

*い。

【0078】適合度統合手段275は第1の適合度推定部272と第2の適合度推定部274から適合度を統合して、情報選択手段24に伝える。適合度統合手段275では2つの適合度推定手段272と274が同じ情報について異なる適合度を推定したときに、適合度の一本化を行なう(A15)。

【0079】情報選択手段24は、適合度統合手段275が一本化した適合度を利用して、要求者の要求に合う情報を選択し(A2)、結果を情報提示部12に出力する(A3)。

【0080】利用者間類似度は従来のSIF方式と同じものが使える。例えば、利用者の各情報に対する評価値を要素としたベクトルを評価ベクトルとして、利用者間の類似度を利用者の評価ベクトル間のピアソンの積率相関係数を使うことが考えられる。

【0081】プロファイル学習方法としては、従来のCBF方式と同じものが使える。ただし、従来は情報の属性と評価値の間の関係を学習していたが、本発明では情報の属性と評価値と情報の属性と第1の適合度推定部が推定した適合度との関係を学習すればよい。例えばプロファイルの学習結果は式3で計算できる。式3中の、評価値もしくは推定適合度、とは評価があれば評価値を、評価が無ければ推定適合度を用いることを示している。

【0082】

【数3】

$$\text{プロファイル} = \frac{\sum_{k=1}^n \text{正規化された} k\text{番目の情報に対する評価値もしくは推定適合度} \times k\text{番目の情報の特徴ベクトル}}{\sum_{k=1}^n \text{正規化された} k\text{番目の情報に対する評価値もしくは推定適合度}} \quad [\text{式3}]$$

【0083】第2の適合度推定部で用いる適合度の推定方法は従来のCBF方式と同じものが使える。例えば情報の特徴ベクトルとプロファイルとのピアソンの積率相関係数を使うことが考えられる。

【0084】第2の適合度推定部では、第1の適合度推定部で適合度が推定できない情報や、推定できても精度が低い情報について、適合度を推定すれば良い。例えば、利用者が誰も評価していない情報や、評価した利用者人数が少ない情報について適合度を推定すれば良い。

【0085】情報統合手段275で2つの適合度推定手段272と274が同じ情報について異なる適合度を推定したときに、適合度の一本化を行なう方法としては、適合度の高い方を選ぶ、適合度の低い方を選ぶ、2つの適合度の平均を取る、適合度の信頼性の高い方を選ぶ等が考えられる。これらの方法から情報の属する分野に合わせて、一本化した後の適合度の信頼性が高くなるように、方法を選べば良い。

【0086】また、本実施の形態の説明に利用した例はあくまで例であり、本発明はこの例だけに限定されるものではない。さらに、本発明の情報フィルタリング方式

14

をコンピュータが実行可能なプログラムとしてCD-ROMやフロッピーディスクに代表される記録媒体に記録してもよい。

【0087】**【発明の効果】**本発明の第1の効果は、従来のCBF方法の情報に価値があるかは結局人が読んで判断しないといけないという問題を軽減していることである。その理由は、適合度推定に人による情報の価値評価が導入されるからである。

【0088】本発明の第2の効果は、従来のCBF方法の未知の属性値や単語を持つ情報に対してフィルタリング精度が低いという欠点を軽減していることである。その理由は、利用者が評価した情報に無い属性値や未知語であっても、他の利用者が評価した情報に含まれていれば、適合度の推定に利用できるからである。

【0089】本発明の第3の効果は従来のSIF方法の問題である誰かが評価・推薦した情報しか得られないことと、ある程度の量の評価が集まらなければフィルタリングの精度が低いことを解決していることである。その理由は、情報の属性も利用して適合度を推定しているた

めである。

【図面の簡単な説明】

【図1】第1の発明の構成の一実施の形態を示すブロック図である。

【図2】第1の発明の動作の一実施の形態を示すフローチャートである。

【図3】第1の発明の適合度推定手段の一構成例を示すブロック図である。

【図4】第1の発明の適合度推定手段の動作の一例を示すフローチャートである。

【図5】従来のCBF方式の構成を示すブロック図である。

【図6】従来のCBF方式の動作を示すフローチャートである。

【図7】従来のSIF方式の構成を示すブロック図である。

【図8】従来のSIF方式の動作を示すフローチャートである。

【図9】従来の技術および本発明が対象とする情報を説明するための図である。

【符号の説明】

1 入出力装置

* 2 データ処理装置

3 記憶装置

4 データ処理装置

5 記憶装置

6 データ処理装置

7 記憶装置

8 データ処理装置

9 記憶装置

11 要求入力部

12 情報提示部

13 評価入力部

21 属性抽出手段

22 利用者プロファイル学習手段

23 適合度推定手段

24 情報選択手段

25 利用者間類似度算出手段

26 適合度推定手段

27 適合度推定手段

271 利用者間類似度算出手段

272 適合度推定手段

273 利用者プロファイル学習手段

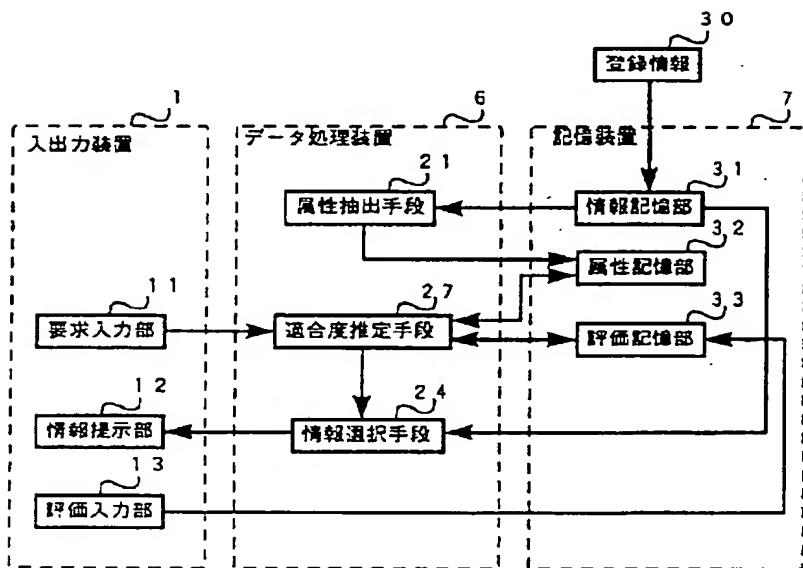
274 適合度推定手段

10

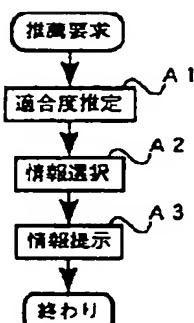
20

* *

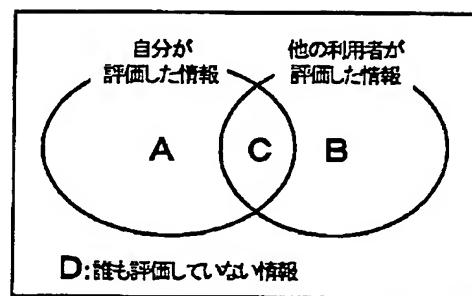
【図1】



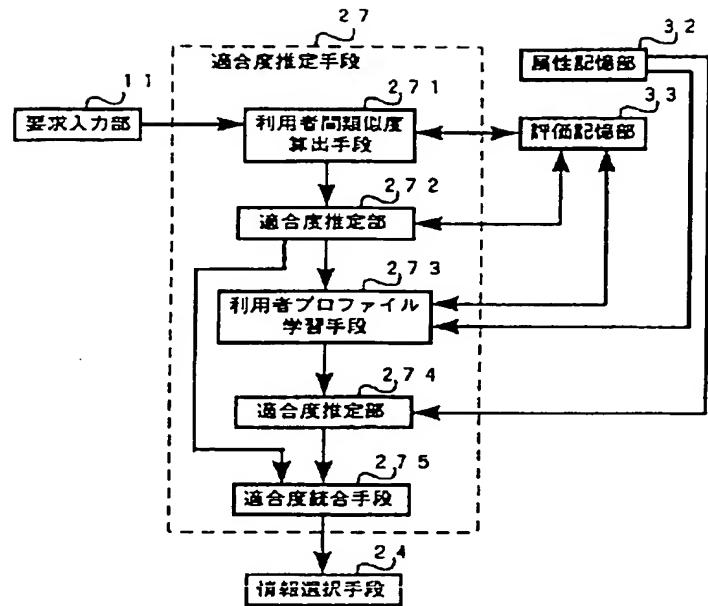
【図2】



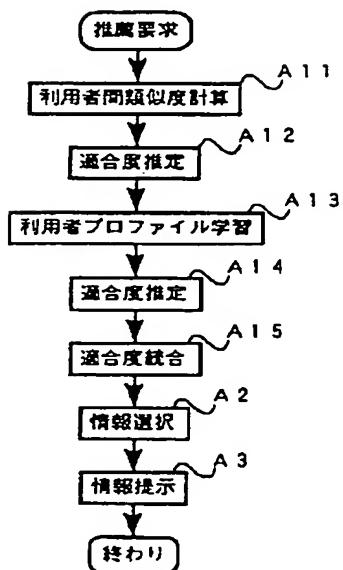
【図9】



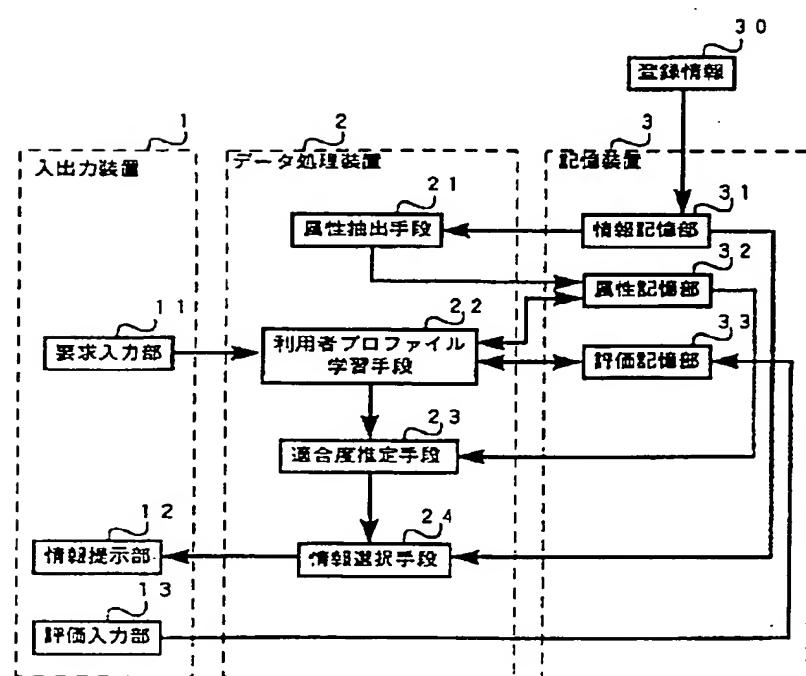
【図3】



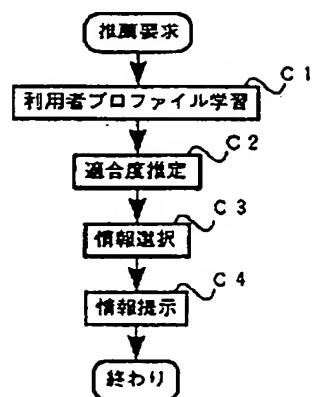
【図4】



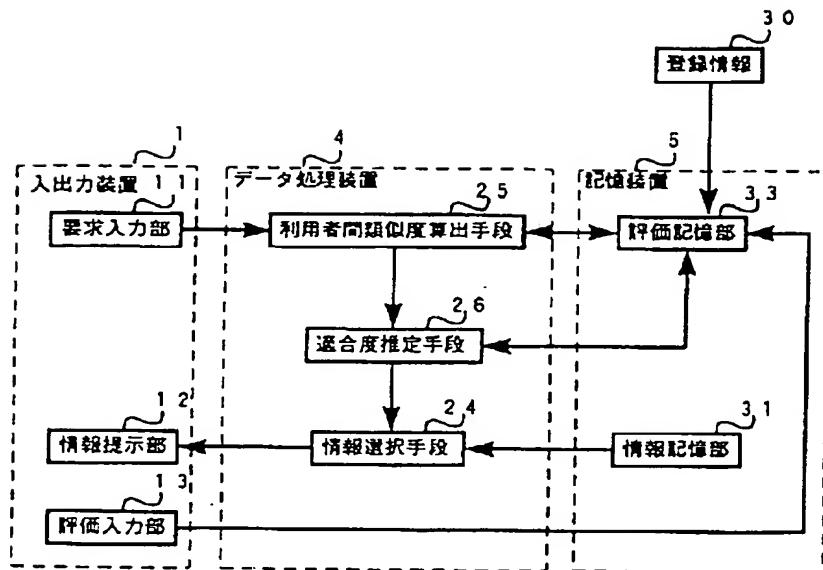
【図5】



【図6】



【図7】



【図8】

